beautiful mind
IAS STUDY GROUP

Research Methods

PAPER- 1 UNIT-3

Major steps in Psychological research (problem statement, hypothesis formulation, research designs, sampling, tools of data collection, analysis and interpretation and report writing) Fundamental versus applied research; Methods of data collection (interview, observation, questionnaire); Research designs (ex-post facto and experimental); Application of statistical technique (t-test, two way ANOVA correlation, regression and factor analysis); Item response theory.

www.beautifulmindias.com

BEAUTIFUL MIND IAS STUDY GROUP

2013

SHOUT
SHARE
IT

Contents

A. Definition of Research 6

A.1 Characteristics of Research 6

A.2 What is Research Methodology? 7

A.3 What is a Psychological Research? 7

A.4 What is Research Process? 7

B. Major Steps in Psychological Research 7

B.1 Problem Statement 7

B.2 Formulating Hypothesis 9

B.2.1 Null Hypothesis (H0) 9

B.3 Research Design 10

B.3.1 Data collection 10

B.3.2 Data Analysis 10

B.3.3 Interpretation 11

B.3.4 Solution and Conclusion 11

B.4 CRITERIA FOR GOOD RESEARCH 11

C. Research Design and Sampling Design 12

C.1 Meaning of research design: 12

C.2 The concept of controlling variance: 13

C.2.1 Procedure for controlling variance: 13

C.3 Characteristics of a Good Research Design 15

D. Sampling 16

D.1 Meaning: 16

D.2 Importance of sampling: 16

D.3 Characteristics of a good sample: 16

D.4 Methods of Sampling 16

D.4.1 Probability Sampling: 16

D.4.2 Non-Probability sampling: 18

E. Non-Experimental, Descriptive or Correlational Research 19

F. Experimental Methods and Quasi Experimental Research 19

F.1 Experimental and control group 20

F.2 Cross sectional method of research 21

F.3 Longitudinal Method 22

F.4 Exploratory case study or single-subject experimental research 22

G. Methods of data collection (TOOLS OF DATA COLLECTION) 23
G.1 Observation 24
G.1.1 Characteristics of Good observation: 25
G.1.2 Advantages of observation 25
G.1.3 Disadvantage of the Method 25
G.2 Checklist 25
G.3 Rating Scales 25
G.4 Questionnaire 26
G.4.1 How to administer Questionnaires? 27
G.5 Interview 28
G.5.1 The merits of interview 29
G.5.2 Limitations of Interview 29
G.6 CRITERIA OF GOOD RESEARCH TOOL 29
G.6.1 Reliability 30
G.6.2 Validity 30
G.6.3 Test of Practicality 31
H. Statistics 31
H.1 Student's t-test 35
H.2 Two-Way ANOVA 35
H.2.1 Assumptions 35
H.2.2 Hypotheses 35
H.2.3 Factors 36
H.2.4 Treatment Groups 36
H.2.5 Main Effect 36
H.2.6 Interaction Effect 36
H.2.7 Within Variation 36
H.2.8 F-Tests 37
H.2.9 Two-Way ANOVA Table 37
H.2.10 Summary 37
H.3 Item response theory 38
H.4 The Item Characteristic Curve 39
H.5 IRT Models 39
H.6 Information 40
H.7 Scoring 40
I. Regression analysis 40
I.1 Underlying assumptions 41

I.2 Linear regression 41

A. Definition of Research

Craford writers, "Research is simply a systematic and refined techniques of thinking, employing specialized tools, instruments and procedure in order to obtain a more adequate solution of a problem that would be possible under ordinary means. It starts with problem, collects data or facts, analyses them critically and reaches decisions based on the actual evidence. It involves original work instead of mere exercise of personal opinion. It involves from a genuine desire to know rather than desire to prove something. It is quantitative, seeking to know not only what, but how much and measurement is therefore a central feature of it".

In fact, research in considered to be a formal, systematic, intensive process of carrying on the scientific method of analysis. It invoices a more systematic structure of investigating, usually resulting in some sort of formal record of procedures and results or conclusions. A scientific research is the search for valid and reliable answer to a problem.

Research has proved to be an essential and powerful tool in leading man towards progress. There would have been very little progress without systematic research. Now we have developed good and useful research and only in psychology but also in the fields like computer. Research is nothing but a matter of raising question and them trying to find answer.

The sociologists, the anthropologists, economists, educationists and psychologists carry on research in their respective field. Each year new products, new facts, new concepts and new ways of doing things came into our lives due to the ever increasing significant research in physical, biological, social and psychologists fields. All significant research leads to progress in some fields of life or the other. So, the goal of all research is progress and development for good and comfortable life.

A.1 Characteristics of Research

Research is a direct method to the solution of a problem. Research used to find out the relationship between two or more variables or variability. It gathers new knowledge and sometimes it replaces the old knowledge.

Characteristics of scientific research are-

- It is characterized by carefully designed and predefined procedure.
- It is based on empirical evidences – i.e. finding answer to the problem is based on data which is observable and measurable. They are not based on intuition.
- It is logical and rational – a good research probes the question from various different angles and each step followed has a logical connection with the other and the result.
- It is impersonal i.e. free from personal opinions
- It is falsifiable – i.e.it is open for criticism and further testing and verification. No research is complete in itself. It can be repeated so that the conclusion derived could be proved or disproved.

Research is a search for knowledge, it is rediscovering existing knowledge and it adds to the existing fund of knowledge of theory. Knowledge enables man to understand, explain, control, and hence cope with any situation. This means the human mind will always search knowledge. It is true yesterday, today, and tomorrow. It endeavors to data in quantitative terms. The objectives of any research are studied under two headings, namely, (a) academic and (b) utilitarian. Research helps in decision making and action.

A.2 What is Research Methodology?

Science of method is methodology. Research Methodology is a science of studying how research is done scientifically. It is a way to solve the research problem systematically. When we talk about research methodology, we not only talk of the research methods but also consider the logic behind the methods.

A.3 What is a Psychological Research?

It is the process of applying the scientific method to the solution of psychological problem or any problems of life in general. The method of obtaining objective solutions to problem encountered in the theory or practice of psychology. The process of arriving at empirically verifiable principles or generalizations is at the root of psychological research.

A.4 What is Research Process?

Research process is a vigorous, impersonal mode of procedure dictated by the demands of logic and objectivity. It is systematic, logical, empirical and replicable. It involves various steps. The various steps involved in a research process are not mutually exclusive nor are they separate and distinct. They do not necessarily follow each other in any specific order and the researcher has to be constantly anticipating at each step in the research process requirements of the subsequent steps. In short, the research process in nothing but a scientific inquiry.

B. Major Steps in Psychological Research

Any research follows 3 stages namely; 1. deciding **what** to research on 2. planning **how** to do it, and 3. actually **doing** the research.

The research processes/steps involved in these stages are as below:

1. problem statement - identification and statement of problem
2. hypothesis formulation
3. research designs
4. data collection
5. analysis and interpretation and
6. report writing

The accuracy of research process depends on carefully and continuously carrying out the various steps.

Such steps/processes overlap continuously rather than following a strictly prescribed sequence. At times, the first step determines the nature of the last step to be undertaken. If subsequent procedures have not been taken into account in the early steps, serious difficulties may arise which may even prevent the completion of the study or research. Every researcher should remember that the various steps involved in a research process are not mutually exclusive; nor are they separate and distinct. They do not necessarily follow each other in any specific order and the research has to be constantly anticipating at each step in the research process the requirements of the subsequent steps.

B.1 Problem Statement

An individual or a group of persons can be said to have either a need felt problem or a problem which can be technically described as a research problem. Microscopically, they (individuals or group) have one or more desired outcomes which are confronted with two or more course of action that have same

but not equal efficiency for the desired objective or objectives and are in doubt about which course of action is best.

Identifying and selecting research problem is a great problem for every researcher. The research must single out the problem he wants to study. He has to select the general area of interest or aspect of subject matter that he wants to study. Initially the problem may be stated in a broad general way and then it should be modified leaving ambiguities and irrelevant things. Then the possibility and feasibility of a particular solution have to be considered before selecting the problem. The selection and formulation of a general topic into a specific research problem is essential.

So then what are the characteristics of a good research problem? A good problem must be

- specific
- well defined
- clear
- empirical
- Testable

This shows the inherent limitation of scientific method- it cannot answer every problem. It can only answer only those problems which can be empirically and logically testable. E.g. mystical problems like "Is there life after death? Does God exist?" - cannot be probed by scientific method and tools available to us right now.

What are the sources of problems? Sources of problems are varied - everyday experiences, academic areas of interests, question banks could provide a research problem.

In formulating the research problem the following steps are to be followed:-

To understand the problem thoroughly, it is good to discuss it with one's own guide, colleagues, and an expert who has research problems in mind. The researchers should review the available related literature of two-types- one the conceptual literature concerning the concepts and theories and another one is the empirical literature consisting of studies made earlier, which are similar to the proposed study. Literature survey would make sure that the problem has not been investigated and proved earlier; even if investigated, how far the results have been conclusive and also helps in delimiting the area of focus of research.

The research problem has to narrow it down and phrase the problem in operational term and rephrasing the same into meaningful terms from an analytical point of view. The statement of the objective is fundamental one because it determines the data which are to be collected and the techniques to be used. Above all, types of problem should also be considered in selection of the research problem.

Review of the Related Literature: It includes facts, concepts, theories, and previous research findings and it is a part of research process. The researcher should undertake the survey of literature related to the problem because it is an eye-opener for research work. Academic journals, conference proceedings, government reports, books, published or unpublished these should be studied, depending on the nature of the study. It is a source for research work for determining sample size, choosing tools and procedure for collecting data, selecting appropriate statistical tools and analyzing and interpreting results. It should be remembered that one source will lead to another. Thus the study of related literature is very useful research work.

B.2 Formulating Hypothesis

First you have to know the meaning of hypothesis. It is beliefs, tentative explanations or assumptions made in order to draw out and test its logical or empirical consequences. It provides focal point for research. It directs researcher in collecting data and analyzing them (in testing the hypothesis). It delimits the area of research and keeps the research on the right track. Formulating hypothesis for any particular problems, requires the following steps:-

a) Understand the objectives in seeking a solution.

b) Review of similar studies on the same problem (literature survey). Discuss the problem with colleagues and experts to formulate hypothesis.

c) Take original field interviews on a limited scale with interested individuals in order to secure greater insight into the real and practical aspects of the problem.

d) Frame the hypothesis in precise and clearly defined terms. (operationalization)

With carefully following above mentioned steps, one can develop or frame or formulate hypothesis, according to the nature of the problem. This is the basic step of the research process.

A good hypothesis possesses the following characteristics-

- It should be in operational terms – i.e. it states the variables that are to identified and measured
- It should be specific
- It should be clear and testable

Finally you have to remember one thing that occasionally you may encounter, some where we do not need framing hypothesis, especially in the case of exploratory or formative researches which do not aim at testing hypothesis.

B.2.1 Null Hypothesis (H0)

A null hypothesis is the logical negative form of the original hypothesis (also known as Alternative hypothesis Ha). When used, the null hypothesis is presumed true until scientific evidence, in the form of a hypothesis test, indicates otherwise — that is, when the researcher has a certain degree of confidence, usually 95% to 99%, that the data does not support the null hypothesis.

H0 = negative form (Ha)

- If you do an experiment to see if a medicine works, the null hypothesis is that it doesn't work.
- If you do an experiment to see if people like chocolate or vanilla ice-cream better, the null hypothesis is that people like them equally.

A researcher, being a human being, is motivated to prove his hypothesis right. This invariably brings in confirmation bias. One way to eliminate this bias is by use of null hypothesis testing.

If scientific observations contradict the prediction of the null hypothesis, it means that either the null hypothesis is false, or the event under observation occurs very improbable. This gives us high confidence in the falsehood of the null hypothesis, which can be improved in proportion to the number of trials conducted. Continued falsehood of null hypothesis makes the original hypothesis to be more valid.

Hence, falsehood of null hypothesis from every possible angle will increase the degree of acceptance of result and level of significance; as confirmation bias of the researcher are controlled.

B.3 Research Design

After having formulated hypothesis, you need to know the condition for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure.

Here, you will have to state the conceptual structure within which research should be conducted. The preparation of such design will provide for the collecting of relevant data or information or evidence with minimal expenditure of effort, time, money, and materials. The effective research design depends mainly on the research purpose. The research purpose may be grouped into four categories, according to the nature of the problem.

They are: (1) exploration (2) description (3) diagnosis and (4) experimentation. When the purpose happens to be an accurate description of a situation or of association between variables, the suitable design will be done that minimizes bias and maximizes the reliability of the data collected and analyzed. Research design includes sampling design, observational design (design or the tools or instruments), operational design and statistical design.

A sample design is a definite plan for obtaining a sample from a given population. It refers to the techniques or the procedure, the researcher would adopt in selecting items for the sample or the size of the sample.

An observational design is a preparation of unstructured or structured instruments tools for the collection of data.

An operational design is advanced decisions about operational procedures.

Statistical design is preplanned design for analysis of the data.

The designs are flexible and different according to the nature of the problem, hypothesis formulated and the purpose of the research. There are many research designs and the researcher must decide in advance as to which design would prove to be more suitable for his research work.

B.3.1 Data collection

It is very important step in the research process. According to research design, you are expected to collect data using reliable and valid tools or instruments, selecting required size of the sample and executing correct procedure. Sometime, the researcher would get training for collection of data and coding the data. Another very important thing that you have to remember is that the data to be collected should be adequate and dependable, and only then the result would be dependable and reliable. With the help of experts and guides, rigorous, and vigorous efforts can be made for collecting adequate and dependable data.

B.3.2 Data Analysis

The collected data are raw in nature and therefore they are called as raw data. They should necessarily be condensed into a few meaningful and manageable groups and classified into some purposeful and usable categories.

The categories of data are transformed into symbols that may be tabulated and counted and it is called as a coding operation. Editing is the procedure that improves the quality of the data for coding and then classified data are put in the form of tables which are called as ‘tabulations’. The tabulation is very

useful is analyzing the data. Further analysis of work can be done by computation by applying various well defined statistical formulae. Hypothesis testing is the next step in the process of analysis. The relationships or differences between or among variables can be tested by using coefficient of correlation and t-test in statistics respectively. In hypothesis-testing though the use of t-test, we can establish whether the difference between two groups is a real one or is the result of random fluctuations. If the difference happens to be real, the inference or the result will be that two samples came from different universes (populations). If the difference is due to chance, the inference or result will be that two samples belong to the same universe (population). Similarly, other statistical tests can be used for testing the hypothesis according to their requirements.

The hypotheses may be tested through the use of one or more statistical tests, depending upon the nature and objectives of research work. Hypothesis-testing will result in either rejecting the hypothesis or in accepting it and it helps up to arrive at certain generalizations and conclusion. Therefore, analysis of data will be helpful for the interpretation of data.

B.3.3 Interpretation

After having tested the hypothesis, the investigator may arrive at generalization to build a theory. If the investigator had no hypothesis to start with, he might seek to explain his findings on the basis of previous studies done by some investigators in the same way or on the basis of some facts and theories. It is known as interpretation between present study and previous studies or findings which may support or against the present investigation or research work.

William Emory says, “In one sense, interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses”.

In short, the interpretation is comparing of present investigation with previous studies or findings done by some researchers in earlier times. The interpretation of the results is the original work of the present investigator and it helps to arrive at generalization. The generalization in turn provides solution to the problem.

B.3.4 Solution and Conclusion

From the interpretation, we can arrive at a generalization which leads to solution to the problem.

The important findings are then discussed, interpreted and conclusions are drawn from the whole study. The findings are statement of factual information based upon the statistical analysis of data. From these findings, conclusions are drawn and stated in relation to the specific objectives of the problem and the proposed hypotheses of the study. Generally, the conclusion will be the essence of the whole study. The conclusion is nothing but the solutions to the problem.

B.4 CRITERIA FOR GOOD RESEARCH

The criteria for good research are:

1. The purpose or goal of the research should be clearly defined.
2. The simple and common concepts should be used.
3. The procedural design of research should be clearly planned and designed to get objective results as much as possible.
4. According to the nature of the problem the sampling design should be used.

5. The tools used for research should have reliability and validity and so that data will have reliability and validity.
6. The research procedure should be defined clearly and sufficiently in detail so that another researcher can repeat the same procedure for verification and advancement. This will maintain the continuity in research.
7. The research should have appropriate method of analysis for collected data.
8. The analysis of data should be sufficiently adequate to reveal its significance from the results:
9. One conclusion should be drawn from the result and discussions because it is an essence of the research work.
10. The prepared research report should be without any bias, fervour and favour.

These are all ten important general criteria for good research. In addition to these, there are four indispensable or special qualities or criteria of good research. They are:

The good research is systematic

It is because that it rules out the use of guessing and intuition in arriving at conclusion or generalization. It is structured with specified steps to be taken in specified sequence in accordance with the well-defined set or rules.

Good research is logical

In fact logical reasoning makes research more meaningful in the context of arriving at conclusion or decision making. So, the rule of logical reasoning is a vital part in the process of research.

Good research is empirical

It deals with one or more aspects of real situation with concrete data. Further it provides a basis for external validity to research result. So, we can say that the research deals with empirical evidence.

Good research is replicable

The repetition is allowed in the research because it is a social science. The research results can be verified by means of replicating the research. This will enable the researcher to take sound and strong decision, or conclusions or generalizations.

Finally you can say that good research is systematic, logical, empirical and replicable, which are very important criteria for research.

C. Research Design and Sampling Design

C.1 Meaning of research design:

Research design is a plan or strategy for conducting the research. Research designs are specific to the types of research. Research is a serious enquiry into a problem. When research is conducted according to a well-designed structure it is most likely to be accurate and purposeful. Research is the process of attempting to answer researchable questions. A research design is a more general plan and that a single design may contain and require the use of several methods. **Kerlinger** (1973) has identified two basic purposes of research design: (1) to provide answers to research questions and (2) to control variance. Research design sets up the framework for adequate tests of relations among variables. Research problem suggests how many observations are to be made, which variables are active and

which are attributes. A design tells us what type of statistical analysis to use. Finally an adequate design outlines possible conclusions to be drawn from statistical analysis. The research design ensures that a researcher obtains usable results.

C.2 The concept of controlling variance:

All research is conducted for the purpose of explaining variance-the fact that all individuals are not the same or have the same score. Also when the variance of any one variable is considered, it may be influenced by any number of factors. Variance in achievement, for example may be due to aptitude and motivation etc. The manifestation of variance can be explained by an example. A high school teacher is studying the effects of three different methods of motivation on achievement in physics. The research problem is stated as follows: "A study of the effects of motivation on the performance of high school students". The problem implies that an experiment should be conducted since motivational training an independent should be conducted since motivational training an independent variable of primary interest will be manipulated by the experimenter. Motivational training has three levels say M1, M2, and M3. The dependent variable is performance in the school subjects, physics which can be measured through achievement tests administered after training. Ninety students enrolled in the same school and taught by the same teacher are tested on achievement. There will be 90 scores. There scores will not be the same and hence the distribution will have variance and it may be due to different effects and since motivation is the independent variable, the researcher has to determine whether or not they have different effects. Some students may be more able to do then others and other physical and mental conditions may cause validity. Therefore it is important to study the effects of variables on the dependent variable. Controlling variance means being able to explain what is causing it.

C.2.1 Procedure for controlling variance:

There are four ways by which controlling variance can be done.

a) Randomization.
b) Building conditions on factors into the design as independent variables.
c) Holding conditions or factors constant.
d) Statistical adjustments.

(**a**) **Randomization**: In the above example the same person gives training to 90 students, having three groups of 30 students each, each group trained by different motivational method. Assume that the 90 students form a representative sample of some larger population, but they are a heterogeneous group with respect to ability level is randomly distributed to three groups and we expect its effect to be the same in each group of 30 students.

It can also be noted that in this example other variables associated with the students for examples anxiety is distributed randomly among the three groups. Mathematical knowledge related to performance in physics is also randomly distributed.

(b) **Building in factors as independent variables**: Of the 90 students, if the top 45 students are classified as 'Higher' and the remaining 45 as 'lower' ability level, then we will have ability level as independent variable with two levels. Fifteen students of each of the two ability groupings will be randomly assigned to each method.

Research Design with ability level built in as an independent variable

Method Ability	**Students Randomly assigned**	**Dependent Variables**

M_1	High	15	Performance
	Low	15	
M_2	High	15	
	Low	15	
M_3	High	15	
	Low	15	

(c) Holding factors constant: Holding a factor constant consists of reducing a variable to a constant. For example, the trainer could reduce ability level to a constant if only students with one defined ability say those scoring between 98 and 108 on an I.Q. test are included in the study. If ability level does not affect performance on the dependent variable, its effect would now be diminished. Only students within a particular I.Q. range would be randomly selected. This number will be less than 30 per group. Here the results arrived generalize only to the restricted group.

(d) Statistical control: Statistical control is attained through computational procedures. When the data are analysed, the variable of ability level in physics assumes that an ability measure consisting of performance on I.Q. test is available for each of the 90 students. There may exist a relationship between I.Q. test scores and performance on the physics test such as high scores on one go with the high scores on the other. If physics test scores are adjusted ability on the physics test scores. This can be done using a sophisticated statistics and the specific statistics procedure commonly used is the analysis of covariance using procedures for control in combination. The four procedures for enhancing control can be used singly or in combination with each other. The four variables (in addition to the independent variable, i.e., motivation) and their methods of control are as follows.

Variable	**Method of Control**
Science background of the student.	Randomization
Trainer	Build in as an independent variable. The same trainer trains everyone in the sample. The same teacher teaches physics.
School	Reduced to a constant. Students of only one school are included.
Ability level	Statistical control. The I.Q. test scores are based.

The purpose of controlling variance is to enhance the interpretation of results so that the researcher can tell what effects, if any, the variables are having. When the research design is developed, the researcher has to think about controlling variables. Careful designing is essential to enhance the validity of the results.

C.3 Characteristics of a Good Research Design

A good research design must not only be appropriate but must yield results from the research that can be interpreted with confidence. Four characteristics of a good research design are identified. However they cannot be considered naturally independent of each other. They are, free from bias, free from confounding, control of extraneous variables and statistical precision for testing hypotheses.

a. Free from bias: This means that the data and the statistics computed from them do not vary in any systematic way except as expected on the basis of random fluctuations. The differences that appear can be attributed to the independent variables under study. Bias can enter into the data in a number of ways, including biased assignment of the individuals to the experimental method.

b. Freedom from confounding: A good research design should eliminate confounding of variables or keeps it to a minimum so that we can separate effects and we can interpret results without confusion. Two or more variables are confounded if their effects cannot be separated. In the physics example, if there are three trainers instead of one, trainer and type have been confounded. If the students trained in one type of motivation score higher on the physics exam than those of the other two types, we will not know whether the higher performance is due to one type or to the trainer. The effects of trainer and type of motivation cannot be separated because each trainer uses only a single type.

c. Control of extraneous variables: the control of extraneous variables means that the influences independent variables outside to the purposes of the study are minimized, nullified or isolated. In other-words, the variance of such variables is in effect reduced to zero or near zero. There are three ways to control extraneous variables. For example if we select only one sex for an experiment, then we can be sure that sex cannot be contributing independent variables. Here we lose the power of generalization. For instance, if we use only boys in the experiment, we cannot say anything about the relation understudy with girls.

The second method to control extraneous variance is through randomization. Theoretically randomization is the only method of controlling all possible extraneous variables. This does not mean that the group should be equal in all the possible variables. By chance, the groups can be unequal but the probability of their being equal is greater.

The third means of controlling an extraneous variable is to build it into the design as an independent variable. For example, if sex is to be controlled, then it can be added as another independent variable.

The fourth way to control extraneous variance is to match subjects. The basic principle is to split a variable into two or more parts. (example high & low intelligence) and then randomize within each level, but care should be taken on the subjects which are matched. They should be related to the dependent variable.

Statistical Precision for testing hypothesis:

Another characteristic of a good research design is that it will provide appropriate data with enough precision to test those hypotheses that require statistical tests. In statistical sense, precision is increased as random (or error) variance is decreased.

D. Sampling

D.1 Meaning:

Sampling may be defined as "the selection of part of an aggregate or totality on the basis of which a judgment or inference about the aggregate or totality is made". A sample is a smaller representation of a large whole. It is section of the population selected from the latter in such a way that they are representative of the universe. A single member of a population is called elements are selected with the intention of finding out something about the population from which they are drawn, that group of elements is referred as a sample and the process of selection of the elements for a sample is called sampling.

D.2 Importance of sampling:

When the population is very large, the research study can satisfactory cover the entire population through sampling. If the number is limited, intensive and exhaustive data collection is possible. It saves time, cost and energy. When the units of area are homogeneous, sampling technique is highly useful.

D.3 Characteristics of a good sample:

It should be representative of the population. It should be free from error due to bias, deliberate selection. It should have adequate size. A randomly selected sample of fairly adequate size, reflecting all the characteristics of the population is considered to be a good sample.

D.4 Methods of Sampling

Sampling techniques may be classified into two broad categories and they are probability sampling and non-probability sampling.

D.4.1 Probability Sampling:

This is a strategy in which sample elements are automatically selected by some method under which every elements has the probability of being selected equally.

a) Simple Random Sample:

Theoretically, the best way of achieving probability sampling is by random method. One example of random sampling is the lottery which is commonly used for selecting prize winners. Here the selection is random or unpredictable and therefore fair because every item has an equal chance to be selected. The selection of one item does not depend upon the other item.

Merits: there is no possibility of personal bias of the researcher in the selection of items. Probability theory can be used for generalization.

Demerits: It is time consuming and expensive. If the items are not homogeneous either in size or in nature, this method cannot be applied.

b) Systematic Sampling:

To simplify the simple random sampling process, systematic sampling is adopted. In this case, the population is finite or accurate. It is estimated by drawing every n^{th} element from an existing list beginning at a randomly determined point. (k = N/n rounded of to the nearest number N = number of elements proposed to constitute the sample, k = 100/5=20. The first element will be selected from the 1^{st} 20 elements, then every 20^{th} element from that will be automatically included in the sample

(example: 6,26,46,66,86,……. etc.) In systematic sampling, only the 1st element is selected randomly and the others are at fixed intervals.

Merits: It is easier and less expensive. The sample is fairly representative of the population.

Demerits: If there is a hidden periodicity in the population, systematic sampling will not be effective. For example if every 20th item is defective, then we would get all defective and the generalization of the result will be questionable.

c) Stratified Random sampling:

In some cases, the population to be sampled is not homogenous but in essence, consists of several sub populations. For example, the population of working women can be divided into higher income group, middle income group, and lawer income group based on their income. A sample is drawn from each stratum of random. These samples are then combined to from a single sample of the universe. The sample so drown will be typical of the whole population as it will represent all the different segment. This approach to sampling is called stratified random sampling because the population is stratified into its sub-population. All the strata are represented in sample.

Merits: All the significant groups are represented. Units are concentrated and localized within each stratum.

Demerits: Bias may be caused in the sample through improper stratification. There are three forms of stratifies random sampling namely, equal, proportional and optimum.

Type	Characteristics
Equal	An equal number of case are taken from each stratum regardless of the size of the strata in proportion to the universe.
Proportional	The number of elements to be draw from each stratum is the same proportion as they constitute in the universe.
Optimum	In cases where starta differ not only is size but also in variability, the larger and more variable the stratum, the greater will be the contribution to the sample. The sampling fraction varies among starta.

d) Cluster sampling:

When the selection of individual members from the population for the sample is impractical or too expensive, groups or clusters of members may be selected for the sample. Cluster sampling is a procedure of selection in which the unit of selection, called the cluster, contains two or more population members. Each member of the population must be uniquely contained in one and only one cluster. In stratified random sampling, selection occurs with individual members but in cluster sampling selection occurs not with individual members but with the clusters. The cluster are randomly selected from the large population of clusters and once a clusters, is selected for the sample of the population

members in that cluster are included in the sample. It is not necessary that all cluster have the same number of population members. In cluster sampling, the exact sample size may not be known until the sample is selected. This is because clusters are not of the same size. For example fourth grade children's achievement motivation is to be estimated using a standardized test. It there are, say 135 elementary schools it will be too expensive to administer the test to all students and a selection by simple random will be extensive. Startified random sampling may be possible but it has one disadvantage. They are in the class and it will be inconvenient to test some members leaving others. Suppose there are 597 fourth grade classes throughout the system with an average enrolment of 34.8 students per class. A sample size of 520 students is desired so it may be decided to select 15 classes or clusters.

Merits: It is economical, easier to observe cluster of units and because it combines the advantages of both random as well as stratified sampling.

Defects: It may produce larger sampling error. Sample bias may be there due to unequal size of the sub groups selected.

Sample stages:

Single stage samples are those drawn in one step from a sampling frame. Multistage samples are drawn in a sequence of two or more samples stages and the last will be elements to considered. Fist we may sample the inclusive groups from the population. Secondly the subgroup will be sampled form the sampled inclusive groups. At the third stage we may select the individual cases from the selected subgroups. In the example cited for cluster sampling, the population is the 4^{th} grade students of the different elementary schools. First we make a random sampling of the required number of schools from the total number of schools (which is the sampling frame of the population). Then from, the selected schools (inclusive groups of the population), we make a random sampling of classes. The classes are the subgroups. Form the subgroups, we then randomly select the individuals, the schools student thus selected form the final sample, first sampling is done at school level, then at the class level and finally at the student level. At every level sampling is done at random. It can be done either simple or stratified at every stage.

D.4.2 Non-Probability sampling:

The alternative to probability sampling is non-probability sampling, which includes any method in which the elements have unequal chances of being selected. There are four kinds of non-probability sampling.

- Purposive Sampling
- Convenience Sampling
- Quota Sampling
- Incidental Sampling.

Purpose Sampling: In this, respondents are chosen because of certain characteristics. For example, career women is suspected to have a risk factor of certain stress disorders but it is rare in the population. In this case, the researcher may resort of purposive sampling by interviewing the particular career women.

Convenience Sampling: It depends on the convenient availability of the respondents. No pre planning is necessary for this. A convenience sample may be used for pilot study. Further convenience sampling may be used in one or more of the following situations.

- When the universe is not clearly defined.
- Sampling unit is not clear

- A complete source list is not available.

Quota Sampling: In this, the researcher tries to create a sample which matches the demographic profile in the population. For example, if it is know that 54 percent of the adult of a community are females and the researcher wants 100 respondents, the first 54 female, and the first 46 males, will be selected by quota to match the population profile. Quota sampling is a simple case of convenience sampling. For example, selecting the 100 respondents from those persons doing business at a shopping center.

Incidental Sampling: Without applying any method described above, whatever comes on our way we take it as a sample. Travelers pickup opinion of the people whom they meet on their way, to write a book. An inspector of school puts a question to a student in a class to understand whether the teaching-learning is upto the mark in a class of students. In the T.V. Channels you may observe interview (in a cinema theatre) of people after seeing a movie whether it is good. Incidental sampling is a casual choice of a sample for observation.

E. Non-Experimental, Descriptive or Correlational Research

A descriptive research describer interprets facts as they are the conditions that exist, the relationship between two events, options that are held, the processes that are going on, effects that are evident, and the trends that are developing. Descriptive research involves events that have already taken place and may be related to present condition and also primarily what is happening now. Hence, they are called ex-post fact i.e., fact after it has happened. For example, there was an evaluation of voting behaviour of people in India by different research groups.

In descriptive research, the investigator or researcher does manipulate events. This method is widely used in behavioural science, because there are many human behaviours which cannot be manipulated, and it is also unethical. For example, to study the effect of deprivation of food, a researcher should not deprive children from eating when they are hungry, to understand its effect on lungs. Thus, in studying such relationship, the researcher has to use descriptive methods where human behaviour can be systematically analyzed and examined in natural settings.

However, descriptive studies also have the criteria of science, namely (a) disciplined enquiry, (b) expertise in the field, (c) objectivity and (d) careful execution.

F. Experimental Methods and Quasi Experimental Research

Experimental method of research involves a logical and systematic way of answering questions. Experimenters, researchers or investigator manipulate certain facts or variables. They control certain conditions in the field and study and observe how the behaviour changes or is affected.

Hence, in experimental research there is a deliberate and systematic manipulation. The researcher must be aware of all the relevant factors that could influence the outcome. Experimentation is the classic method of science. It is more sophisticated, exacting and powerful method of disconcerting and developing an organized body of knowledge.

Great achievement in the field of medicine, science and technology is attributed to this experimentation process.

Experimental method in also used in other fields such as education, agriculture, and also social sciences.

F.1 Experimental and control group

An experiment involves the comparison of the effects of a particular treatment with that of different treatment or no treatment. In a simple conventional experiment, reference usually made to an experimental group.

For example, let us take an example of an experimental study wherein a researcher is interested in knowing the effects of monetary reward on performance of a specific task in a group of school boys. The experiment divide the boys into two groups. These two group are equated as nearly as possible for factors such as age, education, social status, parent's status etc. The experimental group is exposed to the influence of the factor under consideration, i.e., the boys in the experimental group are given monetary reward for performing certain tasks, where as the boys in the control group are not given any monetary reward. Observations on performances and other aspects are systematically recorded by the researcher to determine what difference appear in performance of experimental group as contracted with the control group. Sometimes, the effect of investigator participating in the experiment itself may influence the results. When an experimenter in dealing with physical and chemical objects, the chances of his/her presence influencing the outcome is minimal. However, in dealing with animals and human subjects, the experimenter's variable (presence) may affect the results. This effect in clinical research is called 'placebo effect'.

Medical researchers have long recognized that patients who receive any medication, regardless of its real efficacy, tend to feel better refer form more effectively. In medical experiments, a harmless or inert substitute is administered to the control group to offset the psychological effect of medication. These substitutes, or placebos, are indistinguishable from the real medication under investigation, and neither experimental nor control subjects know whether they are receiving placebo or medication. The effectiveness of true medication is the difference between the effect of the medication and that of the placebo.

Similarly in psychological researchers, the participation in a study influences the behaviour of people over and above the actual treatment given. This effect in called "Hawthorne effect." This was recognized by "Hawthorne" in a series of experiments at the western electric company in 1950s. Just the way medical researchers uses a placebo control group, the psychological researchers use a comparison control group with minimum exposure to experimental set up. Varying types, amounts or degrees of experimental factor may be applied to members of control group.

In the above experiment, the researcher may use a single experimental group with monetary reward and 2 or 3 control groups. One control group may have the presence of experimenter while performing the task and another control may receive a verbal reward or appreciation. Control may be performing the task without any one's presence (no experimentor's presence). There are few concepts we have to understand in relation to the experimental method of research. They are variables, validity and designs. In this unit we will discuss the concepts of variables and experimental validity.

Variables are the conditions or characteristics that the experimenter manipulates, controls and observes for example, age, gender, economic status, attention span, performance, intelligence, achievement etc. **Independent Variables** are the conditions or the characteristics that experimenter manipulates, controls or varies to ascertain its relationship to observed phenomenon. **Dependent variables** are the conditions or characteristics that appear, disappear, or changes as the experimenter introduces, removes or changes independent variables. Independent variable can be treatment and organisic (attribute) variables. Treatment variable are those factors that experimenters manipulate and to which subjects are assigned. Attribute variables are those which cannot be altered and are already

determined, but the experimenter can include or remove them from experiment by manipulating (controls).

Let us see an experimental design to understand the different variables in "a study on efficacy of relaxation on headache activity among women." Here the experimenter is using a treatment variable i.e., relaxation (technique). This he can manipulate by exposing one group to its influence and another group without it. The dependent variable in headache activity or headache experiences on which the effect in seen. As the intensity of number of relaxation increases, the headache activity also should vary. Thus, the study aims to show the cause effect relation between two set of variables.

There are other attribute variable such as gender which may also influence headache experience. Hence sex in controlled by the experimenter by including only women group in the study.

Confounding variables are those aspects of study or sample that might influence the dependent variables and whose effects may be confused with the effects of independent variable.

Confounding variables can be intervening or extraneous. In the above example there are many aspects of a person's life which may have an effect on the headache experience, for example, stress, general health etc. They must be observed and controlled so that they do not influences the outcome (dependent) variable.

Extraneous variables are those which are present during experimental situation and may affect the independent variable or dependent variable experimenter's bias, enthusiasm, testing procedure, etc.

A Good experimental study would control the influence of intervening and extraneous variables on the outcome measures. These aspects you would have studied in the previous lessons.

F.2 Cross sectional method of research

To make significant contribution to the development of knowledge and its application, the researcher must conduct studies which are scientifically valid. The factor that have been manipulated in the study must have definite effect on the observed consequences.

For Example, a teacher who is varying the teaching method (independent variable) must be able observe the true differences on the learning or performance in children (dependent variables) who have received different teaching methods.

But in experimental situations there may be many variables which are present in the situation or generated during experimental situations which may influence the dependent variables.

Let us examine some of them now

School children change biologically and psychologically in many ways over a period of time. This is called "Maturation Effect" which may influence the learning and performance rather than the teaching method (independent variable).

Specific events in the child's life (called history) may also affect the performance and learning (For example, anxiety of exam, family problems etc.)

Testing procedures and instruments vary sometimes and they may influence the results.

In order to control such influences, researcher can resort to cross sectional studies or longitudinal study.

A cross sectional method involves studying a large sample of subjects across different age groups. For example, the teacher who is studying the effect of teaching method may take sample of school children from different class levels 7^{th} standard, 8^{th} standard and 9^{th} standard. He or she may take a representative sample of the subjects in each group or class. Thus, "maturation" effect can be controlled by studying how younger and older children vary in their learning and performance when exposed to different teaching method.

A cross sectional study use multiple sample in different sections or over different period of time.

The advantage of this method are that a large sample of subjects can be studied at a give time. Whatever factors which are influencing the independent variable can be controlled by the investigator using cross section method. For example, age, sex, time, geographical location. It is economical and generalisability of the study is high.

F.3 Longitudinal Method

Studies of development, growth and charge always involve groups using time as a variable. Longitudinal studies involve obscurations by the experimenter at different pints in the time. It is possible to control for the reaction to measurement effects in the longitudinal study.

For example, while testing the children an investigator may notice that performance may be poor because of high reactivity to testing procedure. However, in the longitudinal study, the children will overcome such reactivity and their performance can also vary.

Generally longitudinal studies are used in single group studies. However, controls groups are also used in longitudinal researchers method to control extraneous factor influencing the dependent variables.

Longitudinal method are useful in studying process oriented changes. The internal validity and external validity of the experiments can be very well demonstrated here. Adding more control groups, adding more observation helps in replication of the experimental results and demonstrating the efficiency of experimental interventions. Generally survey research or expost facto type descriptive studies are cross-sectional in nature. Field experiment and explorative interventional studies are longitudinal in nature.

Longitudinal research uses times-series and other multivariate statistical procedures to analyze the information.

Longitudinal research method suffers from experimental mortability of subjects (dropouts from investigation) over exposure. They are time consuming. In spite of these limitations many advances in behavioural science is attributed to the knowledge gathered from decades of studies conducted by behavioural scientists.

F.4 Exploratory case study or single-subject experimental research

The exploratory case study is also referred to as single case or number of one research. This is also a particular type of experimental research. The important feature of this method is the rigorous study of the effect of intervention on an individual. Even though the focus of this type of study is on individual subjects, some of the studies include more one subject. When the number of subject is more, each individual is taken as a separate case and the data is analyzed separately rather than as a group.

Single case study is also a method to test the research hypotheses. One of the major drawback of the case studies is the limitation of generalisability. The evidence based on a single case cannot be generalized to other subjects. Single case studies are widely sued in the area of behaviour modification.

Case study requires careful assessment, repeated observations and measurements and careful control and applications of the experimental treatment.

Observations are critical in studying the cases. The procedure for observation and the behaviours to be observed must be carefully planned. Major psychological process in the field of clinical psychology and developmental psychology is attributed to the single case explorative study. For example, Piaget studied the cognitive development process observing the behaviour of his own children. Many major theories in psychology is based on single case observation. Ebbinghayse's experiments on self were basis for his expositions on memory process. Freud's psychoanalytic theory was based on single case clinical studies.

The single case studies provide preliminary hypotheses for developing major theory. And they are also called explorative studies. There do not have formalized hypotheses for testing, which can be later structured.

Many children and Diagnostic research follow the case study methods. These methods use in depth approaches to understand basic causal relation and also study the processes of change or growth. Observed are carefully defined and are observable. The measurement procedures including tests, surveys, attitude scales are also used. Still the observation is the primary measurement tool in case studies.

In group research designs, what we call pretest data is considered baseline in case studies. Thus, observations are made under different conditions such as pre-intervention phase, intervention phase etc. There are different designs used in case studies which you would have studied in previous lessons on research designs.

G. Methods of data collection (TOOLS OF DATA COLLECTION)

Once the research problem and plan in chalked out, the researcher has to collect the evidence for testing the hypothesis. The researcher would have to decide which sort of information he/she would be using for the study. According to the decisions they have to use different methods of data collection. Data collection is an important process in research, which is essential in arriving at conclusions. The data should be appropriate and also adequate for the study.

The data can be primary or secondary. The former is collected by the researcher directly either by survey or experimentation. For example, if a researcher conducts an experiment,he can record his observation on a daily record. This forms his primary data. Secondary data are those which are already collected by someone else and which have already been passed through the statistical process.

The method used by the researcher to collect information or evidence for testing hypotheses is called tools. The data can be collected by more than one way be using tools like observation. Interviews, questionnaires, checklists, inventories, rating scale etc.

The tools used by the researcher to collect the data depends on the following factors:

- Nature of investigation
- Objective and scope of inquiry
- Financial resources
- Available time
- Desired degree of accuracy.

Ability to select the appropriate tools for research depends on the experience and knowledge of the researcher.

Let us examine each of the above mentioned tools of data collection to ascertain what ways they are adequate or appropriate for the research.

G.1 Observation

Observation is a common scientific tool for collecting information. Here the information is sought by way investigators, own direct observation without asking the respondent. This method in widely used in behavioural science.

For example instead asking the respondent what brand of pen he/she uses, the investigator may look at the pen he uses. Many preference of people can be inferred from observation.

Observation:- Observation of natural phenomena along with systematic classification and measurement lead to the development of various theories and laws. Observation is a part of any king of research, be it an experimental one or a descriptive type. Observation method is mostly used in single subject experimental research as it is a costly affair to observe behaviours pertaining to a larger group. Observation can be direct data gathering method in descriptive research. This data is know as anecdotal reports and very useful fo research studies.

Observation must always be systematic, for a specific purpose, carefully focused and thoroughly recorded. Like any other research procedures, it must be subjected to the usual formalities of accuracy, validity and reliability. It is very necessary that the observer knows that what is significant or insignificant to the study. Objectivity, being careful and accurate in measuring and recording important aspect of observation.

Since human behaviour is complex and certain tracts are difficult to observe directly, they must be carefully used in precise operational form. It is also important that behaviours that mean different things to different observers, be dealt carefully. In defining which behaviour needs to be acted out, the researches needs to first determine the class roles avoid labeling permissible behavior as acting out. Instruments like stopwatch, mechanical counter, camera, audiometer, audio, video recordings and ocher devices help in making observations more sincere when behaviour are recorded. It is at the proposal of the researcher to start and stop the action for mere accurate recording of the data, to collect inter-observer reliability data without more than one researcher at the site and to re-examine ideas and decide on a new format for coding behaviours. Systematic observation of human behaviour in natural settings is to an extent an intrusion into the dynamics of the situation. This intrusion may be reactive i.e., might cause an effect on the behavoiur of the person who is being observed.

Recording Observations :- Recording while the behaviors is in action is advisable if it would not distract on create a barrier between the observer and the person. This minimizes the errors that might occur from faulty memory. Recording must be done as soon as possible when the action is still fresh in memory. But many agree that objectivity is more likely when the interpretation of the meaning of behaviour described is learnt until a later time as simultaneous recording and interpretation often interfere with objectivity.

Systematizing Data Collection procedures :- To record information obtained through observation, devices like checklists, rating scales, scorecards and scaled specimens help in systematic means of summarizing.

G.1.1 Characteristics of Good observation:

Observation is planned carefully, systematic and perspectives. Observers are aware of significant and insignificant details in their research. They aren't distracted by the action or seeing itself. Observers are aware of the wholeness of what they observe. They are objective. Bias which may influence what they observe and report, is eliminated to the maximum. They separate the facts from the interpretation of facts. Interpretation is done later. Observations are checked and verified by repetition or comparison with those of other observers. Observation are carefully and expertly recorded. Observations are collected in such a way to check for reliability and validity.

G.1.2 Advantages of observation

Subjective bias on the part of respondent is eliminated. In other words, what is obvious behaviour cannot be hidden from investigator, if observation is accurate. Information on what is currently happening can be obtained. Respondent's willingness to respond does not count. This method in very useful in studying behaviours of the, children and clinical cases where respondents are not ready to verbalize or not capable of giving verbal responses.

G.1.3 Disadvantage of the Method

It is expensive and time consuming. Information provided is very minimal. Unforeseen factors may interfere with observation.

G.2 Checklist

Checklist is a simple device of data collection comprising of prepared lists of behaviours or items. The presence or absence of a behaviour is indicated by checking 'Yes' or 'No' or the type of number of items may be indicated by inserting the right number. Checklists can also be used to count the number of behaviours occurring in a given time period.

G.3 Rating Scales

This involves qualitative description of a limited number of aspects. The classifications may vary from 5 to 7 point scale of description like.

- Superior/above average/average/fair/inferior
- Always/frequently/occasionally/rarely/never
- Strongly agree/agree/disagree/strongly disagree

Rating scale may be either graphic scale or an itemized rating scale. The above example are all graphic scales where the rater indicates his rating by a simple tick mark at the appropriate point on a line that runs from one extreme to the other.

This type rating may sometime increase the tendency of rater to choose always a particular point (tendency to mark on neutral point). The frame of reference for the researcher and the respondent may vary. This is another limitation.

Itemized rating scale is also known as numerical scale. Here a series of statements is presented and the respondents select the statements that are as best reflecting his view or perception or he may order them according to his preference. The order is progressively in terms of more or less on some property. Here is an example.

What do you prefer most in your job?

- Leisure time
- Conductive work atmosphere

- Good company policy
- More perks
- Responsibility
- Opportunity to use potentials
- Independence
- Good wages
- Good relationship

This type of scale provides more information to the rater and increases its reliability. Rating scales are simple. They have wide range of applications. There are three possible errors in Rating scale. They are (a) error of leniency, (b) error of central tendency and (c) error of hollo effect.

When respondents are too leniently rate items and rate some hardly, it is called error of leniency. They consistently rate strongly agree for almost all the statements.

When the respondents are reluctant to rate, it results in choosing a middle path i.e., neural point or least change. This is called error of central tendency.

The error of follow effect operates when a respondent carries over a generalized bias from one statement to another. For example, a person may rate an individual as beautiful and because of this rating, he may generalize that the person must be smart, intelligent, sociable etc.

One problem in the construction of a rating scale is conveying to the rater exactly which quality one wishes to evaluate. For this to be considered, the traits and categories must be carefully defined in observable terms.

The Limitations of rating scales are as follows:

It is difficult to define that trait or characteristic to be evaluated. It is leads to halo effect in judgment. Rater tends to be too generous. Example: concluding about 60 to 80% are above average in any particular item.

G.4 Questionnaire

Another most important tool used in the research is questionnaires. It is also called as inquiry form. In this data gathering tool, the respondents answer to questions or to statements in writing. When an investigator is interested in factual information, questionnaire are used. Administrations of questionnaires to single individuals or to groups is possible. This way the investigator can establish rapport, explain the purpose of the study, and explain the meaning of items that may not be clear.

Availability of high proportion of data and also consumption of lesser time to gather data makes this tool a popular research instrument for investigators. Even though, the data collection through questionnaires are easier, the construction of questionnaires will demand lot of time, hard-work and ingenuity of the researcher.

Questionnaire that calls for shorts, check mark responses like ‘yes’ or ‘no’ is known as the restricted or closed form type. It may also require to check an item from a list of suggested response. The closed form in easy to fill, takes little time and keeps the responses objective and is fairly easy to tabulate and analyze.

Open form or unrestricted questionnaire calls for a free response in the respondent’s own words. For example, the following item is a questionnaire is an open ended question.

'How long have you been smoking?'

Constructing a questionnaire for the research purpose involves many scientific principles.

Questionnaires must be in written languages and clearly worded. Meanings of the words must be clear and specific. They should not mean different things to different persons. There are many ambiguous words which have to be clarified. Using adjectives like frequently, rarely etc may have different meanings for different individuals. Hence an investigator must specify the frequency very clearly in terms of times per week, or times per month etc.

Do not use double negatives in framing questions. While giving alternative for marking responses it must be clear. To underline a word to indicate special emphasis, the question must be phrased in such a way that it is appropriate for all the respondents.

A Good questionnaire deals with significant topic relevant to research propose. It must be as short as possible and only long enough to get the essential data. Question never which are neatly arranged, clearly printed and easy to respond are well received by the respondents.

The order of questionnaire item must be from general to specific. Annoying and embarrassing questions must be atleast not at the beginning. Following the above principles, many questionnaires are developed by psychologists, social psychologists and educationists to investigate complex aspects of human behaviour. Many psychological tests are in the questionnaire forms. There are achievement tests, aptitude scales, interest schedules, personality questionnaires, and certain semi projective inventories. There are certain scientific principles involved in construction of the questionnaires. The development of tools of measurement is an advanced area of psychological research. This field is also called as 'psychometry' in this lesson, you will study in detail the criterion for establishing a questionnaire as a reliable and valid tool for the research.

G.4.1 How to administer Questionnaires?

A researcher can either construct a questionnaire for the purpose of his/her study. Or can use a questionnaire, which has already been constructed and standardized by another researcher.

When a researcher is using an newly constructed tool, it is essential to try study sample. Trying out the questionnaire on a pilot sample or another group can help in clearing doubles or confusion which a respondent may notice. This trial run is worth the effort, as it reveals many defects which can be corrected before the final format in printed and administered to a large sample.

Thus, trial run in important not only to refine the newly constructed questionnaire, but also to test the receptivity for the questionnaire.

A researcher who is using questionnaire has to choose the respondents very carefully.

The questionnaire must be relevant for the group who are potential respondents. For example, majority of items on occupational stress will be irrelevant to those who are unemployed.

A questionnaire in English will not be appropriate to a respondent who cannot read or write in that language.

Another important factor is that the respondents must be willing to respond. Hence sometimes written consent will be helpful. Cooperation and willingness is essential to get the reliable information on questionnaires.

A researcher should get permission and approval from concerned authorities before administering the questionnaires. For example hospital authorities should permit before one can assess the patients on certain personality test. This holds good for educational institutions, private and public organization and communities also.

As for as possible the investigator must reveal the true intention of the survey, and assure the confidentiality of survey, and also-non-judgmental attitude towards the opinion or attitude or people.

Most of the questionnaires can be administered individually or in group situations. Sometimes the researcher may insist on spontaneous responses. Otherwise the respondent may be given time to reflect and answer. Generally the questionnaires are not time bound, the respondents have their over freedom to answer.

While administering the questionnaire to a large group, an investigator must take precautions on seating arrangements and uniformity of instructions. Assistants, help can be taken to get good and clear response from the subjects.

An incomplete or wrongly coded responses can lead to a great loss of data. Lack of comprehension of instruction can lead to error in marking responses, and individuals may not respond to few items. This leads to great loss of data. Hence investigator must always check for completed questionnaire.

G.5 Interview

Interview in another important tool of investigation for a researcher. The interviews in a method which requires a person known as 'interviewer' asking questions generally in a face to face contact to the other person. Interview is a direct personal investigation. Sometimes, telephone or indirect mass media may be used to gather information on one to one basis.

Interview method in always suitable for intense and confidential investigations. When interview is conducted in an orderly and structured way it is called structured interview. At other times the interviews may be unstructured and informal also. The structured interview has predetermined questions, and helps the investigator in maintaining the direction of the interview. On the other hand, unstructured interview provides opportunity for the investigator to be more free to explore, change the sequence of enquiry or modify the queries according to the situations.

Unstructured interview does not provide standard for comparability of two interview. It is more time consuming. The interviewers must be very skillful and be of high caliber and possess deep knowledge in the subject.

In explorative case studies, diagnostic studies, formulative research studies and specific clinical research method, unstructured interview is used as a valuable tool for gathering research evidences. On the other hand, in descriptive studies where generalization of observations has to be made on a small sample, structured interview is used as it is being more economical and less time consuming.

Unstructured interview can also be classified as focused interview and non-directive interview. Focused interview is meant to focus attention on the given experience of the respondent and its effects. The interviewer will confine the respondent to a discussion of issues which are relevant.

In the non-directive interview, the interviewer will encourage the respondent to talk about the given topic with a bare minimum of direct questioning.

The interviewer will often acts as a 'catalyst' to a comprehensive expression of the respondents, feelings and beliefs. This method in widely used in psychotherapy sessions.

G.5.1 The merits of interview

The researcher gain wide range of information and also of greater depth unlike questionnaire which merely states 'yes' or 'no' to statements. The interview provides the deep feelings, beliefs and other orientation of interviewee to any issue. Interviewer can use his/her personal skills to establish rapport. A questionnaire is too distant an approach. However, interview can break the resistance on the part of respondent and help in gaining cooperation and truthful information. Unlike a questionnaire, interview provide freedom to reframe questions and issues. Observation as a tool can be successfully incorporated along with interviews. Interview help in gaining contact, sustaining the response, getting additional or supplementary information.

G.5.2 Limitations of Interview

It is expensive and time consuming. Approachability of the sample varies for different strata. Some respondents may not be approachable at all. Presence of interviewer may create over enthusiasm in some subjects. Much depends on the interviewer's skills, knowledge and training. The interviewer must be properly selected and trained.

G.6 CRITERIA OF GOOD RESEARCH TOOL

A researcher who is gathering data to test hypothesis must use tool of the investigation appropriately. If the "tool" is wrong then, the evidence cannot be used successfully to test the hypothesis. There are certain important criteria based on which a research tool or instrument is selected for the purpose of study.

They are the reliability and validity of the tools. These two are very essential for effectiveness of data gathering.

Reliability is the degree of consistency that the instrument or procedure demonstrates whatever the tool is measuring. It must measure consistently. Reliability of the data gathered for the research depends on the reliability of the tool used. Whenever a test, questionnaire or inventory is administered repeatedly, it must give consistent results.

Validity is that quality of data gathering instrument which enables it to measure what it is supposed to measure.

Reliability is necessary but not sufficient condition for validity. A tool may be reliable but may not be valid.

It is possible through various statistical treatments to establish quantitatively the reliability and validity of the tools. For example, when a researcher states that the psychological tool has a reliability coefficient of 0.70 then he is quantitatively stating that the information obtained is consistent or reliable to the extent of 70%. Higher the quotient, greeter is the consistency.

Similarly when a researcher states that the diagnostic questionnaire has a predictive validity of 0.85, it implies that a diagnosis of disease in the person based on the questionnaire, and his chance of getting a disease may be correct to the extent of 85%.

G.6.1 Reliability

Reliable tests are stable in whatever they measure and yield comparable scores upon repeated administration. On the other hand, an unreliable test yields different measurement each time it is applied.

The reliability of an instrument can be expressed in terms of correlation coefficient. There are different types of measuring reliability.

Reliability or stability over time: This is also called as test retest reliability. The scores of the subjects correlate high with first and second tests or successive administration of tests.

Reliability over items or statements: some tests have two or more forms. They may correlate highly.

Stability of items: It scores on certain items in the test are highly correlated with scores on other items then there is internal consistency or stability. Different statistical procedures are used to measure the internal consistency of the tool.

Stability over scores: certain psychological tools can be interpreted by two different judges. Two persons can independently scores the same set of test papers. When these two sets of scores are highly correlated then the reliability of the tool is higher.

Standard error of measurement: this is a statistical procedure which permits us to examine whether the given measure is a true measure of a person trait's or qualities. This tells us how much the obtained score differs from the individuals true score. Lesser the difference between the obtained scores and true score, greater is the reliability of measurement (scores obtained on the test. Reliability of the testing instrument can be increased by increasing the member of items of equal quality. Carefully design direction for administration of test. Testing situation must be free from distractions. There should be no boredom and fatigue during testing situation.

G.6.2 Validity

Basic of the validity of a tool is to measure the right thing or asking right questions. The items of a questionnaire, inventory must appropriately sample a significant aspect of the purpose of the investigation.

Sometimes validity is also thought as utility. There are many kinds of validity depending on the purpose of the tool in the research. They are (1) Content Validity, (2) Construct Validity, and (3) Criterion related validity (predictive validity and concurrent validity).

Content validity: This refers to the degree to which the tool (test) actually measures, or is specifically related to the traits for which it was designed. The examples are as follows:

How adequately a test paper measures the knowledge and skills that a students is expected to master in a course of study in a week. How adequately a personality questionnaire purporting to measure jealousy has items which indicate jealous feelings, behaviours or statements. How adequately an interest schedule includes statements which elicit information on subjects, likes and dislikes about a profession or job.

Construct Validity: this is the degree to which scores on a test can be accounted for by the explanatory constructs of a sound theory. Intelligence tests are based on different theories. Hence each intelligence test shows a measure which is appropriate to the theory on which it is based. Construct validity is very important for personality tests for example a test measuring type A personality can be validated by

asking the judges to rate the behaviours and gestures of people who score high on the test. The higher rating of the judges and higher scores are found to establish the construct the “Type A behaviour”.

Predictive validity: Refers to the usefulness of a test in predicting some future performance. For example, a researcher is interested in developing a diagnostic psychology testing tool. For this, he wants to discriminate the person for various diseases. Accordingly, the test score shows high for the diseases persons and scores low for the normal persons. In this case, the test has a high degree of predictive validity.

Concurrent Validity: This refers to the usefulness of a test in closely relating to other measures (or) scores on another test of known validity. Tests are often validated by comparing their results with a test of known validity. For example, Minnesota Multiphasic Personality Inventory (MMPI) has been found to be very widely accepted as a valid tool. Many other personality tests which have been developed later have been validated against MMPI. Many performance tests are validated against paper-pencil tests. Sometimes the researcher can validate the scores on a particular test with that of a rating given by significant relatives of the subjects.

A test may be reliable even though it is not valid. Thus the test can yield scores consistently score consistently scores but of no value. On the other hand, a test which is measuring a specific concept must definitely measure the same concept consistently. This shows reliability and validity are important criteria for a good research tool. There is also another criterion, practicability.

G.6.3 Test of Practicality

The practicality characteristic of a measuring instrument can be judged in terms of economy, convenience and interpretability. From the operational point of view, the measuring instrument ought to be practical i.e., it should be economical, convenient and interpretable.

Economy: Tests that can be given in a short period of time are likely to gain the cooperation of the subjects and to conserve the time of all those involved in test administration. The matter of expense of administrating a test is often a significant factor if the testing program is being operated on a limited budget.

Interest: Tests that are interesting and enjoyable help to gain the cooperation of the subject. Those that are dull or seem silly may discourage or antagonize the subject. Under these unfavourable conditions, the test is not likely to yield useful results.

Interpretation: test scores and result must be amenable for interpretation by the researcher objectively and uniformly.

H. Statistics

"What is statistics?" This is a fair question to ask especially because it is so central to most scientific disciplines today. Statistics is a collection of mathematical techniques that help to analyze and present data. Statistics is also used in associated tasks such as designing experiments and surveys and planning the collection and analysis of data from these.

Statistics can be applied to various different problems and situations but the underlying concepts all remain the same. Thus it is important to understand what statistics is, not only from an application point of view but also from an interpretation point of view. This is required because of the diverse applications of statistics, from social science experiments to studying quantum mechanical phenomena.

Statistics can be broadly classified into descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

H.1 Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.

Different areas of study require different kinds of analysis using descriptive statistics. For example, a physicist studying turbulence in the laboratory needs the average quantities that vary over small intervals of time. The nature of this problem requires that physical quantities be averaged from a host of data collected through the experiment.

Numerical measures are used to tell about features of a set of data. There are a number of items that belong in this portion of statistics, such as:

- The average, or measure of center, consisting of the mean, median, mode or midrange.
- The spread of a data set, which can be measured with the range or standard deviation.
- Overall descriptions of data such as the five number summary.
- Other measurements such as skewness and kurtosis.
- The exploration of relationships and correlation between paired data.
- The presentation of statistical results in graphical form.

H.2 Inferential Statistics

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.

Most predictions of the future and generalizations about a population by studying a smaller sample come under the purview of inferential statistics. Most social sciences experiments deal with studying a small sample population that helps determine how the population in general behaves. By designing the right experiment, the researcher is able to draw conclusions relevant to his study.

While drawing conclusions, one needs to be very careful so as not to draw the wrong or biased conclusions. Even though this appears like a science, there are ways in which one can manipulate studies and results through various means. For example, data dredging is increasingly becoming a problem as computers hold loads of information and it is easy, either intentionally or unintentionally, to use the wrong inferential methods.

There are two major divisions of inferential statistics:

- A confidence interval gives a range of values for an unknown parameter of the population by measuring a statistical sample. This is expressed in terms of an interval and the degree of confidence that the parameter is within the interval.

- Tests of significance or hypothesis testing tests a claim about the population by analyzing a statistical sample. By design there is some uncertainty in this process. This can be expressed in terms of a level of significance.

Both descriptive and inferential statistics go hand in hand and one cannot exist without the other. Good scientific methodology needs to be followed in both these steps of statistical analysis and both these branches of statistics are equally important for a researcher.

Statistics is an independent branch and its use is highly prevalent in all the fields of knowledge. Another way of dividing the branch of statistics is on the characteristics of the data and methods used. Many methods and techniques are used in statistics. These methods have been grouped under parametric and non –parametric statistics.

The field of statistics exists because it is usually impossible to collect data from all individuals of interest (population). Our only solution is to collect data from a subset (sample) of the individuals of interest, but our real desire is to know the "truth" about the population. Quantities such as means, standard deviations and proportions are all important values and are called "parameters" when we are talking about a population.

Since we usually cannot get data from the whole population, we cannot know the values of the parameters for that population. We can, however, calculate estimates of these quantities for our sample. When they are calculated from sample data, these quantities are called "statistics." A statistic estimates a parameter.

H.3 Parametric Statistics

Parametric statistical procedures rely on assumptions about the shape of the distribution (i.e., assume a normal distribution) in the underlying population and about the form or parameters (i.e., means and standard deviations) of the assumed distribution.

These normally involve data expresses in absolute numbers or values rather than ranks.

The parametric statistical tests operates under certain conditions. These consitions are not ordinarily tested and they are assumed to hold valid. The meaningfulness of the results of a parametric test depends on the validity of the assumptions.

The basic assumptions of parametric statistics are:

- The population form which the sample have been drawn should be normally distributed. Normal distribution refers to frequency distribution following a normal curve, which is infinite at both ends.
- The variables involved must have been measured at interval or ratio scale.
- The observation must be independent. The inclusion or exclusion of any case in the sample should not unduly affect the result of the study. The selection of one case in the sample is not dependent on any other case.
- The samples have equal or nearly equal variance. This condition is known as equality or homogeneity of variance and is particularly important to determine when samples are small.

- The population from which the samples are drawn must have same variance, or in some cases must have same ratio of variance. This is called as homosedasticity.

Examples of Parametric tests are t-test, z-test, F-Test and ANOVA

H.4 Non-parametric Statistical tests

To contrast with parametric methods we will define nonparametric methods. There are very few assumptions for non-parametric statistics to apply. These are statistical techniques for which we do not have to make any assumption of normality for the population we are studying. Indeed, the methods do not have any dependence on the population of interest. It is for this reason that nonparametric methods are also referred to as distribution free methods.

We face many situations where we cannot meet the assumptions and conditions and thus cannot us parametric statistical procedures. In such situation we are bound to apply non-parametric statistics.

If our sample is in the form of nominal or ordinal scale and the distribution of sample is not normally distributed, and also the sample size is very small, it is always advisable to make use of the non-parametric tests for comparing samples and to make inferences or test the significance or trustworthiness of the computed statistics. In other words, the use of non-parametric tests is recommended in the following situations:

1. Where sample size is quite small.
2. When assumptions like normality of the distribution of scores in the population are doubtful
3. When measurement of data is available either in the form of ordinal or nominal scales or when the data can be expressed in the form of ranks and in the shape of + signs or –signs and classifications like 'good-bad" etc
4. Nature of the population from which samples are drawn is not known to be normal

Nonparametric methods are growing in popularity and influence for a number of reasons. The main reason is that we are not constrained to making as many assumptions about the population that we are working with as what we have to make with a parametric method.

Analysis Type	Example	Parametric Procedure	Nonparametric Procedure
Compare means between two distinct/independent groups	Is the mean systolic blood pressure (at baseline) for patients assigned to placebo different from the mean for patients assigned to the treatment group?	Two-sample t-test	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Was there a significant change in systolic blood pressure between baseline and the six-month follow-up measurement in the treatment group?	Paired t-test	Wilcoxon signed-rank test
Compare means between three or more distinct/independent groups	If our experiment had three groups (e.g., placebo, new drug #1, new drug #2), we might want to know whether the mean systolic blood pressure at baseline differed among the three groups?	Analysis of variance (ANOVA)	Kruskal-Wallis test
Estimate the degree of association between two quantitative variables	Is systolic blood pressure associated with the patient's age?	Pearson coefficient of correlation	Spearman's rank correlation

I. Student's t-test

in statistics, a method of testing hypotheses about the mean of a small sample drawn from a normally distributed population when the population standard deviation is unknown.

In 1908 William Sealy Gosset, an Englishman publishing under the pseudonym Student, developed the *t*-test and *t* distribution. The *t* distribution is a family of curves in which the number of degrees of freedom (the number of independent observations in the sample minus one) specifies a particular curve. As the sample size (and thus the degrees of freedom) increases, the *t* distribution approaches the bell shape of the standard normal distribution. In practice, for tests involving the mean of a sample of size greater than 30, the normal distribution is usually applied.

It is usual first to formulate a null hypothesis, which states that there is no effective difference between the observed sample mean and the hypothesized or stated population mean—i.e., that any measured difference is due only to chance. In an agricultural study, for example, the null hypothesis could be that an application of fertilizer has had no effect on crop yield, and an experiment would be performed to test whether it has increased the harvest. In general, a *t*-test may be either two-sided (also termed two-tailed), stating simply that the means are not equivalent, or one-sided, specifying whether the observed mean is larger or smaller than the hypothesized mean. The test statistic *t* is then calculated. If the observed *t*-statistic is more extreme than the critical value determined by the appropriate reference distribution, the null hypothesis is rejected. The appropriate reference distribution for the *t*-statistic is the *t* distribution. The critical value depends on the significance level of the test (the probability of erroneously rejecting the null hypothesis).

For example, suppose a researcher wishes to test the hypothesis that a sample of size $n = 25$ with mean $x = 79$ and standard deviation $s = 10$ was drawn at random from a population with mean $\mu = 75$ and unknown standard deviation. Using the formula for the *t*-statistic, the calculated *t* equals 2. For a two-sided test at a common level of significance $\alpha = 0.05$, the critical values from the *t* distribution on 24 degrees of freedom are −2.064 and 2.064. The calculated *t* does not exceed these values, hence the null hypothesis cannot be rejected with 95 percent confidence. (The confidence level is $1 - \alpha$.)

A second application of the *t* distribution tests the hypothesis that two independent random samples have the same mean. The *t* distribution can also be used to construct confidence intervals for the true mean of a population (the first application) or for the difference between t

I.1 Two-Way ANOVA

The two-way analysis of variance is an extension to the one-way analysis of variance. There are two independent variables (hence the name two-way).

I.1.1 Assumptions

- The populations from which the samples were obtained must be normally or approximately normally distributed.
- The samples must be independent.
- The variances of the populations must be equal.
- The groups must have the same sample size.

I.1.2 Hypotheses

There are three sets of hypothesis with the two-way ANOVA.

The null hypotheses for each of the sets are given below.

- The population means of the first factor are equal. This is like the one-way ANOVA for the row factor.
- The population means of the second factor are equal. This is like the one-way ANOVA for the column factor.
- There is no interaction between the two factors. This is similar to performing a test for independence with contingency tables.

I.1.3 Factors

The two independent variables in a two-way ANOVA are called factors. The idea is that there are two variables, factors, which affect the dependent variable. Each factor will have two or more levels within it, and the degrees of freedom for each factor is one less than the number of levels.

I.1.4 Treatment Groups

Treatment Groups are formed by making all possible combinations of the two factors. For example, if the first factor has 3 levels and the second factor has 2 levels, then there will be 3x2=6 different treatment groups.

As an example, let's assume we're planting corn. The type of seed and type of fertilizer are the two factors we're considering in this example. This example has 15 treatment groups. There are 3-1=2 degrees of freedom for the type of seed, and 5-1=4 degrees of freedom for the type of fertilizer. There are 2*4 = 8 degrees of freedom for the interaction between the type of seed and type of fertilizer.

The data that actually appears in the table are samples. In this case, 2 samples from each treatment group were taken.

	Fert I	Fert II	Fert III	Fert IV	Fert V
Seed A-402	106, 110	95, 100	94, 107	103, 104	100, 102
Seed B-894	110, 112	98, 99	100, 101	108, 112	105, 107
Seed C-952	94, 97	86, 87	98, 99	99, 101	94, 98

I.1.5 Main Effect

The main effect involves the independent variables one at a time. The interaction is ignored for this part. Just the rows or just the columns are used, not mixed. This is the part which is similar to the one-way analysis of variance. Each of the variances calculated to analyze the main effects are like the between variances

I.1.6 Interaction Effect

The interaction effect is the effect that one factor has on the other factor. The degrees of freedom here is the product of the two degrees of freedom for each factor.

I.1.7 Within Variation

The Within variation is the sum of squares within each treatment group. You have one less than the sample size (remember all treatment groups must have the same sample size for a two-way ANOVA) for each treatment group. The total number of treatment groups is the product of the number of levels for each factor. The within variance is the within variation divided by its degrees of freedom.

The within group is also called the error.

I.1.8 F-Tests

There is an F-test for each of the hypotheses, and the F-test is the mean square for each main effect and the interaction effect divided by the within variance. The numerator degrees of freedom come from each effect, and the denominator degrees of freedom is the degrees of freedom for the within variance in each case.

I.1.9 Two-Way ANOVA Table

It is assumed that main effect A has a levels (and A = a-1 df), main effect B has b levels (and B = b-1 df), n is the sample size of each treatment, and N = abn is the total sample size. Notice the overall degrees of freedom is once again one less than the total sample size.

Source	**SS**	**df**	**MS**	**F**
Main Effect A	*given*	A, a-1	SS / df	MS(A) / MS(W)
Main Effect B	*given*	B, b-1	SS / df	MS(B) / MS(W)
Interaction Effect	*given*	A*B, (a-1)(b-1)	SS / df	MS(A*B) / MS(W)
Within	*given*	N - ab, ab(n-1)	SS / df	
Total	sum of others	N - 1, abn - 1		

I.1.10 Summary

The following results are calculated using the Quattro Pro spreadsheet. It provides the p-value and the critical values are for alpha = 0.05.

Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F-crit***
Seed	512.8667	2	256.4333	28.283	0.000008	3.682
Fertilizer	449.4667	4	112.3667	12.393	0.000119	3.056
Interaction	143.1333	8	17.8917	1.973	0.122090	2.641
Within	136.0000	15	9.0667			

Total	**1241.4667**	**29**				

From the above results, we can see that the main effects are both significant, but the interaction between them isn't. That is, the types of seed aren't all equal, and the types of fertilizer aren't all equal, but the type of seed doesn't interact with the type of fertilizer.

I.2 Item response theory

In psychometrics, **item response theory** (IRT) is a body of theory describing the application of mathematical models to data from questionnaires and tests as a basis for measuring abilities, attitudes, or other variables. It is used for statistical analysis and development of assessments, often for high stakes tests such as the Graduate Record Examination. At its most basic level, it is based on the idea that the probability of getting an item correct is a function of a latent trait or ability. For example, a person with higher intelligence would be more likely to correctly respond to a given item on an intelligence test.

Formally, IRT models apply mathematical functions that specify the probability of a discrete outcome, such as a correct response to an item, in terms of person and item parameters. Person parameters may, for example, represent the ability of a student or the strength of a person's attitude. Item parameters include difficulty (location), discrimination (slope or correlation), and pseudoguessing (lower asymptote). Items may be questions that have incorrect and correct responses, statements on questionnaires that allow respondents to indicate level of agreement, or patient symptoms scored present/absent.

Among other things IRT theory provides a basis for evaluating how well assessments work, and how well individual questions on assessments work. In education, Psychometricians apply IRT in order to achieve tasks such as developing and refining exams, maintaining banks of items for exams, and equating for the difficulties of successive versions of exams (for example, to allow comparisons between results over time).

IRT is often referred to as *latent trait theory*, *strong true score theory*, or *modern mental test theory* and is distinguished from Classical test theory.

properties of a given assessment, and therefore also the quality of estimates.

From the perspective of more traditional approaches, such as classical test theory, an **advantage** of IRT is that it potentially provides information that enables a researcher to improve the reliability of an assessment. This is achieved through the extraction of more sophisticated information regarding psychometric properties of individual assessment items. IRT is sometimes referred to using the word *strong* as in *strong true score theory* or *modern* as in *modern mental test theory* because IRT is a more recent body of theory and makes more explicit the hypotheses that are implicit within Classical test theory.

IRT models are often referred to as *latent trait models*. The term *latent* is used to emphasise that discrete item responses are taken to be *observable manifestations* of hypothesized trait, construct, or attribute, not directly observed, but which must be inferred from the manifest responses. Latent trait models were developed in the field of sociology, but are virtually identitical to IRT models.

I.3 The Item Characteristic Curve

The performance of an item in a test is described by the *item characteristic curve* (ICC). The curve gives the probability that a person with a given ability level will answer the item correctly.Persons with lower ability have less of a chance, while persons with high ability are very likely to answer correctly.

Much of the literature on IRT centers on item response models for the ICC. A given model describes the probability of a correct response to the item as a function of a *person or ability parameter* (or, in the case of multidimensional item response theory, a vector of person parameters). This probability depends on one or more *item parameters* for the *item response function* (IRF). For example, in the *three parameter logistic* (3PL) model, the probability of a correct response to an item *i* is:

$$p_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-a_i(\theta - b_i)}}$$

where θ is the person (ability) parameter and a_i, b_i, and c_i are the item parameters.

The item parameters simply determine the shape of the IRF and in some cases have a direct interpretation. The parameter b_i represents the item location which, in the case of attainment testing, is referred to as the item difficulty. It is point on θ where the IRF has its maximum slope. The example item is of medium difficulty, since b_i=0.0, which is near the center of the distribution. Note that this model scales the item's difficulty and the person's trait onto the same continuum. Thus, it is valid to talk about an item being about as hard as Person A's trait level or of a person's trait level being about the same as Item Y's difficulty, in the sense that successful performance of the task involved with an item reflects a specific level of ability.

The item parameter a_i represents the discrimination of the item: that is, the degree to which the item discriminates between persons in different regions on the latent continuum. This parameter characterizes the slope of the IRF where the slope is at its maximum. The example item has a_i=1.0, which discriminates fairly well; persons with low ability do indeed have a much smaller chance of correctly responding than persons of higher ability.

For items such as multiple choice items, the parameter c_i is used in attempt to account for the effects of guessing on the probability of a correct response. It indicates the probability that very low ability individuals will get this item correct by chance, mathematically represented as a lower asymptote. there is a 1/4 chance of an extremely low ability candidate guessing the correct answer, so the c_i would be approximately 0.25. This assumes that all options are equally plausible, because if one option made no sense, even the lowest ability person would be able to discard it.

The *two parameter logistic* model (2PL) is equivalent to the 3PL model with c_i = 0. The 2PL model is appropriate for testing items where guessing the correct answer is highly unlikely, such as write-in tests.

I.4 IRT Models

Broadly speaking, IRT models can be divided into two families: unidimensional and multidimensional. Unidimensional models require a single trait (ability) dimension θ. Multidimensional IRT models model response data hypothesized to arise from multiple traits. However, because of the greatly increased complexity, the majority of IRT research and applications utilize a unidimensional model.

IRT models can also be categorized based on the number of scored responses. The typical multiple choice item is *dichotomous*; even though there may be four or five options, it is still scored only as correct/incorrect (right/wrong). Another class of models apply to *polytomous* outcomes, where each

response has a different score value. For example, the polytomous Rasch model is a generalisation of the Rasch model that applies to data in two or more ordered categories. A common example of this Likert-type items, e.g., "Rate on a scale of 1 to 5."

I.5 Information

One of the major contributions of item response theory is the extension of the concept of reliability. Traditionally, reliability refers to the precision of measurement (i.e., the degree to which measurement is free of error). And traditionally, it is measured using a single index defined in various ways, such as the ratio of true and observed score variance. This index is helpful in characterizing a test's average reliability, for example in order to compare two tests. But IRT makes it clear that precision is not uniform across the entire range of test scores. Scores at the edges of the test's range, for example, generally have more error associated with them than scores closer to the middle of the range.

Item response theory advances the concept of item and test information to replace reliability. Information is also a *function* of the model parameters. For example, according to Fisher information theory, the item information supplied in the case of the Rasch model for dichotomous response data is simply the probability of a correct response multiplied by the probability of an incorrect response, or,

$$I(\theta) = p_i(\theta)q_i(\theta).$$

The standard error of estimation (SE) is the reciprocal of the test information of at a given trait level, is the

$$\mathrm{SE}(\theta) = 1/\sqrt{I(\theta)}.$$

Thus more information ideal (the *observed score*) is always the true score occluded by some degree of error. This error may push the observed score higher or lower.

Also, nothing about these models refutes human development or improvement. A person may learn skills, knowledimplies less error of measurement.

I.6 Scoring

After the model is fit to data, each person has a theta estimate. This estimate is their score on the exam. This "IRT score" is computed and interpreted in a very different manner as compared to traditional scores like number or percent correct. However, for most tests, the (linear) correlation between the theta estimate and a traditional score is very high (often it is .95 or more). A graph of IRT scores against traditional scores shows an ogive shape implying that the IRT estimates separate individuals at the borders of the range more than in the middle.

It is worth noting the implications of IRT for test-takers. Tests are imprecise tools and the score achieved by an indivge or even so called "test-taking skills" which may translate to a higher true-score.

J. Regression analysis

Regression analysis is a technique used for the modeling and analysis of numerical data consisting of values of a dependent variable (response variable) and of one or more independent variables (explanatory variables). The dependent variable in the *regression equation* is modeled as a function of the independent variables, corresponding parameters ("constants"), and an *error term*. The error term is treated as a *random variable*. It represents unexplained variation in the dependent variable. The

parameters are estimated so as to give a "best fit" of the data. Most commonly the best fit is evaluated by using the least squares method, but other criteria have also been used.

Data modeling can be used without there being any knowledge about the underlying processes that have generated the data; in this case the model is an empirical model. Moreover, in modelling, knowledge of the probability distribution of the errors is not required. *Regression analysis* requires assumptions to be made regarding probability distribution of the errors. Statistical tests are made on the basis of these assumptions. In regression analysis the term "model" embraces both the function used to model the data and the assumptions concerning probability distributions.

Regression can be used for prediction (including forecasting of time-series data), inference, hypothesis testing, and modeling of causal relationships. These uses of regression rely heavily on the underlying assumptions being satisfied. Regression analysis has been criticized as being misused for these purposes in many cases where the appropriate assumptions cannot be verified to hold. One factor contributing to the misuse of regression is that it can take considerably more skill to critique a model than to fit a model

The term "regression" was coined in the nineteenth century to describe a biological phenomenon, namely that the progeny of exceptional individuals tend on average to be less exceptional than their parents and more like their more distant ancestors. Francis Galton, a cousin of Charles Darwin, studied this phenomenon and applied the slightly misleading term "regression towards mediocrity" to it. For Galton, regression had only this biological meaning, but his work was later extended by Udny Yule and Karl Pearson to a more general statistical context. Nowadays the term "regression" is often synonymous with "least squares curve fitting".

J.1 Underlying assumptions

- The sample must be representative of the population for the inference prediction.
- The dependent variable is subject to error. This error is assumed to be a random variable, with a mean of zero. Systematic error may be present but its treatment is outside the scope of regression analysis.
- The independent variable is error-free. If this is not so, modeling should be done using Errors-in-variables model techniques.
- The predictors must be linearly independent, i.e. it must not be possible to express any predictor as a linear combination of the others..
- The errors are uncorrelated, that is, the variance-covariance matrix of the errors is diagonal and each non-zero element is the variance of the error.
- The variance of the error is constant (homoscedasticity). If not, weights should be used.
- The errors follow a normal distribution. If not, the generalized linear model should be used.

J.2 Linear regression

In linear regression, the model specification is that the dependent variable, y_i is a linear combination of the *parameters* (but need not be linear in the *independent variables*). For example, in simple linear regression for modeling *N* data points there is one independent variable: x_i, and two parameters, β_0 and β_1:

straight line: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \ \ i = 1, N$

In multiple linear regression, there are several independent variables or functions of independent variables. For example, adding a term in x_i^2 to the preceding regression gives:

parabola: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, \ \ i = 1, N$

This is still linear regression as although the expression on the right hand side is quadratic in the independent variable x_i, it is linear in the parameters β_0, β_1 and β_2.

In both cases, ε_i is an error term and the subscript *i* indexes a particular observation. Given a random sample from the population, we estimate the population parameters and obtain the sample linear regression model: $y_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i + e_i$The term e_i is the residual, $e_i = y_i - \widehat{y}_i$.

K. Item Response Theory

Since the 1970s, there has been a sharp upsurge of interest in a class of mathermatically sophisticated procedures for scaling the difficulty of test items. Because of the extensive computation required, these procedures became practicable only with the increasing availability of high speed computers. Although differing in complexity and in specific mathematical procedures, these approaches were originally grouped under the general title of *latent trait models.* The basic measure they used is the probability that a person of specified ability (the so called latent trait) succeeds on an **item** of specified difficulty. There is no implication, however that such latent traitsor underlying abilities exist in any physical or physiological sense, not hat they cause behavior. The latent traits are statistical constructs that are derived mathematically from empirically observed relations among test responses. A rough, initial estimate of an examinee's latent trait is the total score he or she obtains on the test. In order to avoid false impression created by the term "latent trait", some of the leading exponents of the procedures have substituted the more precisely descriptive term "item response theory" or IRT.

Essentially IRT models are used to establish a uniform "sample free" scale of measurement that is applicable to individuals and groups of widely varying ability levels and to test content of widely varying difficulty levels. Like the use of fixed reference group for normative interpretation of test performance, IRT models need anchor items or a common test as a bridge across examinee samples and across tests or sets of items. However, rather than using the mean and Standard Deviation (SD) of a specific reference group to define the origin and unit size of the scale, IRT models set origin and unit size in terms of data representing a wide range of ability and item difficulty, which may come from several samples. Usually the origin is set near the center of this range. The common scale unit is derived mathematically from the item data; it has several advantages, both theoretical and practical, over the earlier item analysis procedures.

IRT is gradually being incorporated in large scale testing programs. For example, beginning in 1982, this procedure was adapted for equating total scores on the news forms of SAT, so as to express them on the continuing , uniform scale.

The general problem of test equating, whereby scores from different forms of a test are expressed on a uniform score scale, has received increasing attention recently.